Childbearing and labour force participation in South Africa: sibling composition as an identification strategy?

Gabrielle van der Stoep

University of KwaZulu-Natal

1. December 2008

Childbearing and labour force participation in South Africa: sibling composition as an identification strategy?

Gabrielle van der Stoep

July 2009

Abstract

This paper uses same sex sibling composition as a strategy to identify the exogenous effects of childbearing on female labour force participation in South Africa. International studies typically find that sibling sex composition is strong instrument for childbearing, satisfying the two properties of an instrumental variable. First, it is positively correlated with childbearing because couples have a preference for a mixed or balanced sibling sex composition. Second, sex sibling composition is a random occurrence and therefore exogenous with respect to female labour force participation. In contrast to international studies, this paper provides evidence that same sex sibling composition is a poor instrument for childbearing among Africans in South Africa. Two-stage least squares estimation of the effects of childbearing on female labour force participation reveals the weak explanatory power of sibling sex composition in first-stage estimations. The result is that very large standard errors are generated on instrumental variable estimates which cannot be interpreted with any level of confidence.

1. Introduction

The key problem for investigating a causal relationship between childbearing and labour force participation is the endogeneity of childbearing. The decision to bear children may itself be a function of decisions to participate in the labour force; and unobserved heterogeneity in characteristics may be correlated with both of these decisions. Therefore childbearing variables are likely to be correlated with the error term in an equation of labour force participation. This results in biased ordinary least squares (OLS) estimates of the effects of childbearing on labour force participation.

Instrumental variable (IV) estimation has typically been used in the literature in dealing with endogeneity. In order for IV estimation to yield consistent and unbiased estimates, the exogenous instruments chosen must satisfy two properties. The first property, or exclusion restriction, requires that the instrument be uncorrelated with omitted or unobserved variables in the error term. The instrument must therefore have no direct association with the outcome variable of the study, in this case labour force participation. The second property requires the instrument to be strongly correlated with the endogenous explanatory variable i.e. childbearing (Wooldridge 2003:486; Bound *et al* 1995:443).

An instrument that has been argued to satisfy these two properties is the sibling sex composition of a woman's first two children. Same sex sibling composition was first suggested by Angrist and Evans (1998) to investigate the causal effects of further childbearing on mothers' labour force participation in the United States. It was later used by Iacovou (2001) for a similar study in the United Kingdom and most recently by Cruces and Galiani (2007) in the developing country context of Argentina and Mexico. Each study argues that the sibling sex composition of parents' first two children is expected to be correlated with further childbearing because parents generally have a preference for a mixed sibling sex composition or a 'taste for balance' (Ben-Porath and Welch 1976). Compared to parents with a boy and girl, parents whose first two children are of the same sex will be more likely to have a third birth in hopes of attaining a mixed sibling sex composition. But the sex of any given child is argued to be a random occurrence (Clark

2000:95) uncorrelated with background characteristics or unobservable characteristics that influence women's labour force participation. As a result sibling sex composition is expected to be uncorrelated with omitted variables in the error term.

In this study, I investigate whether same sex sibling composition is an appropriate identification strategy to estimate the exogenous effects of childbearing on labour force participation among African women aged 20 to 49 in South Africa. Despite evidence of a strong correlation between sibling sex composition and further childbearing in the United States, the United Kingdom, Argentina and Mexico, this observed correlation is weak among Africans in South Africa. Same sex sibling composition therefore fails the second property of an IV. The implication is that very large standard errors are obtained on IV estimates and coefficients lack precision and any meaningful interpretation (Wooldridge 2003). A possible reason for the weak correlation is that family sizes are larger among Africans in South Africa than among women in the aforementioned countries (Central Intelligence Agency 2008; Ben-Porath and Welch 1976). Although parents may have a 'taste for balance', they may not be concerned about whether their first two children are of the same sex if they desire large families; they eventually expect to obtain a mixed sibling sex composition (*ibid*:292).

In the next section I present some background information. Section three discusses the data and samples to be used. Section four outlines the methodology. Section five evaluates whether sibling sex composition variables meet the two properties of an IV for childbearing in South Africa. Finally, I present second-stage results from the two-stage least squares estimation in section six and then conclude the chapter.

2. Background

Generally empirical studies on the relationship between childbearing and labour force participation find a negative relationship, controlling for a range of covariates (Boushey 2008; Chapman *et al* 2001; Angrist and Evans 1998; Gormick *et al* 1996; Bronars and

Grogger 1994; Browning 1992). The strength of this relationship is sensitive to whether or not childbearing is treated as endogenous within labour force estimations. Typically the strength of this negative relationship is overestimated when childbearing is treated as exogenous within labour force estimations. A key explanation for this result is that mothers may be different from non-mothers in terms of their unobserved characteristics. For example, mothers may have lower levels of ambition or ability, lowering their probability of labour force participation. This indirect effect will bias upward the true direct effect of childbearing on labour force participation using ordinary least squares regression (Aguero and Marks 2008).

To avoid endogeneity bias, IV estimation has been used to disentangle the causal mechanism between childbearing and labour force participation. In earlier studies, instruments for childbearing have included the occurrence of twin births (Bronars and Grogger 1994; Rosenzwieg and Wolpin 1980) and religious affiliation (Ryder and Westoff 1972 cited in Iacovou 2001). Applying twin births as an instrument for childbearing requires very large samples unavailable in many countries, including South Africa, because only very few women report multiple births in household surveys (Connelly *et al* 2006). Using religious affiliation is typically limited to contexts with strong ties to Catholicism.

Sibling sex composition of a woman's first two children (Cruces and Galiani 2007; Iacovou 2001 and Angrist and Evans 1998) is a less data-demanding instrument. However, it can only be used to instrument for *further* childbearing among women with *at least* two children. Sibling sex composition is unrelated to whether a woman shifts her fertility behaviour from zero children to one child; therefore it is not suitable for determining the effects of having any children, as opposed to no children at all, on women's labour force participation (Aguero and Marks 2008; Vere 2008:304, 312). Vere (2008:304-5) notes that "IV estimates of the marginal effects of higher order births are still conditional on a woman's decision to have a first child…if there are economies of scale to be realised in raising children, the marginal effect of the first birth may be much higher than the effects of subsequent births."

Very recent studies in developing countries have identified IV strategies to determine the marginal effect of the first birth (or whether or not a woman has any children at all). Aguero and Marks (2008) use a woman's fecundity or infertility status to instrument for motherhood status in six Latin American countries. Vere (2008) exploits a phenomenon of higher fertility levels in more auspicious years of the Chinese Lunar calendar to identify the marginal effect of a first birth on women's labour force participation in Hong Kong. However application of the Chinese Lunar calendar is limited to countries with dense Chinese populations; and data on infertility status is not available in nationally representative household surveys in South Africa.

3. Data and definitions

Information on the sibling sex composition of children born to a woman *is* determinable using the General Household Survey (GHS) 2002 - a nationally representative household survey of approximately 30,000 households in South Africa. Included among detailed retrospective fertility data in the GHS 2002 is the birth order and sex of each child born alive to women between the ages of 12 and 50. It is possible therefore to determine whether the first two children born to a woman are of the same or different sex.

Unfortunately same sex sibling composition can only be used to instrument for whether a woman has *more* than two children rather than any children at all. Mothers are identified here as women who have had *at least two live births* and at least one child is still alive and 18 years or younger.[1] A final sample of 7,477 African mothers aged 20 to 49 is used in this study.

In the international literature of the effects of childbearing on labour force participation, a co-residency requirement is typically placed on the definition of motherhood. This is a logical restriction because children are not expected to constrain the labour force

[1] I also exclude from the sample of mothers women who are currently in school because they do not constitute potential labour force participants.

participation of mothers who are not co-resident with their children. However in a context of high levels of female labour migration and child fosterage, Posel and van der Stoep (2008) find that a sizeable sample of African mothers in South Africa is not co-resident with their children. These not co-resident mothers are more likely than all other women to be labour force participants. The remaining sample of co-resident mothers are a non-random sample of mothers who are less likely than all other women to participate in the labour force. A co-residency restriction on motherhood therefore exacerbates sources of endogeneity, biasing upward the negative relationship between motherhood and labour force participation (Posel and van der Stoep 2008).

To avoid this additional source of endogeneity bias, I distinguish between both co-resident and not co-resident mothers where 13 percent (990/7,477) of the total sample of mothers (having had at least two births) is not co-resident with their children. A co-resident mother is a woman living with at least one biological child aged 18 years or younger in the same household. A not co-resident mother is a woman with at least one biological child aged 18 years or younger but who does not live with any of her children. This distinction is made using a question in the birth module which asked the mother if she lived with her children in the household. This information is then corroborated using a maternal relationship question in the household roster which asked all household members to identify his/her mother if she lived in the same household.

I have also further disaggregated the sample by marital status where almost 60 percent (4,389/7,477) of the total sample of mothers with at least two children is married. Compared to single mothers there may be stronger desires among couples for a mixed sibling sex composition. Alternatively, higher fertility rates among married mothers compared to unmarried mothers may lower concerns about the sex composition of their first two children because they eventually expect to obtain a balanced sibling composition. I identify a woman as married if her response to a question on marital status was “married or living together as husband and wife".[2] Unmarried women include women who have never married or are divorced, separated or widowed.

4. Methodology

For the sample described above I estimate the following specification:

$$Y_i = \alpha + \beta T_i + \delta X_i + e_i \qquad (1)$$

T_i is the key variable which takes on a value of one if a woman has more than two live births and zero if she only has had two live births. Y_i represents the labour force participation of a mother and is equal to one if a woman is a labour force participant and zero otherwise. X_i is a vector of other explanatory variables including a woman's age, age squared, age at first birth, her educational status and marital status. Household compositional variables, dummy variables for province of residence and whether the woman lives in a rural or urban area are also included among the explanatory variables.

In this estimation, β is supposed to measure the causal relationship between having a third birth and women's labour force participation; however its estimation is complicated by the endogeneity of childbearing. Due to the presence of unobserved heterogeneity and a possible two-way causal relationship between childbearing and labour force participation, T_i may be correlated with the error term, e_i (Aguero and Marks 2008; Bound *et al* 1995).

[2] It is not possible to distinguish women who have had a civil/customary marriage from women who are cohabiting with a partner, but not married, in the GHS 2002. There may be differences in the level of male support and income sharing among couples depending on whether they are married or cohabiting. Furthermore in the event of separation, women in cohabiting partnerships are also less protected by the law and may be more likely to keep their jobs than women in civil marriages as a form or economic security. I acknowledge that these factors may possibly result in differences in labour force participation behaviour by whether women are married or cohabiting with a partner; but I will not be able to observe these differences in this analysis. There may also be heterogeneity in preferences for sibling sex composition among couples who are married compared to cohabiting couples which also cannot be observed here.

Ordinary least squares (OLS) estimation of equation (1) gives biased and inconsistent estimates of β, the effect of a third birth on mother's labour force participation. In particular the magnitude of β is likely to be biased upwards due to unobserved ambition or ability which positively influences the outcome variable but is negatively correlated with having a third birth.

In order to address the endogeneity concern in equation (1) I instrument for *more than two births*, T_i. Let S_i be an indicator equal to one if a woman's first two children are of the same sex or zero if the first two children are of mixed sex. Where a preference exists for balanced families with both girls and boys, *more than two children*, T_i, is expected to be positively correlated with same sex sibling composition (Angrist and Evans 1998; Iacovou 2001; Cruces and Galiani 2007; Arnold 1992; Ben-Porath and Welch 1976). In other words women whose first two children are of the same sex are more likely to have third child than women with a mixed sibling sex composition. But there may be different effects on the likelihood of having a third birth depending on whether the first two siblings of the same sex are girls or boys. For example, the effect may be stronger if the first two children born are girls because there may be a stronger preference for sons than daughters. Therefore, I also decompose same sex siblings into two instruments *two girls*, G_i and *two boys*, B_i which take on a value of one if a woman's first two children born alive were both girls, or both boys, respectively.

Two-stage least squares estimation is then used to obtain the IV estimator of the effect of an exogenous (third) birth on a woman's labour force participation. In the first-stage estimation, the endogenous variable T_i is expressed as a linear function of all other exogenous variables X_i in the structural model and of the identifying instrument/s to obtain the reduced form equation. The first-stage equation relating *more than two children* to sibling sex composition is

$$T_i = \gamma(S_i) + \lambda X_i + \eta_i \qquad (2a)$$

where γ is the first-stage effect of sibling sex composition on the likelihood of having a third birth. Alternatively the first-stage relationship between T_i and sibling sex composition is further decomposed to

$$T_i = \gamma_o(G_i) + \gamma_1(B_i) + \lambda X_i + \eta_i \qquad (2b)$$

In the second-stage the predicted values of T_i are then used in equation (1) as an instrument for whether or not a woman has more than two children.

5. Is same sex sibling composition a good instrumental variable in the context of South Africa?

If same sex sibling composition, S_i, is to be classified as a good instrument it must be i) uncorrelated with the error term, e_i, or exogenous within equation (1) and ii) correlated with whether a woman has more than two births, T_i. If either of the two IV properties is not met then IV estimation can lead to large standard errors, asymptotic bias and inconsistent estimates (Wooldridge 2003; Bound *et al* 1995). For this reason I first test and discuss whether same sex sibling composition satisfies these two properties and is thus a good instrumental variable for further childbearing among African mothers in South Africa.

Previous studies have shown that same sex sibling composition satisfies the two properties of an IV in the United States, the United Kingdom, Argentina and Mexico (Angrist and Evans 1998; Iacovou 2001; Cruces and Galiani 2007). But same sex sibling composition may not be a good instrument for childbearing in South Africa. In particular, it is likely to fail the second IV property.

Evidence from studies on sibling sex preferences does not suggest a preference for a mixed sibling sex composition in South Africa. Using questions asked on desires for additional children in demographic health surveys from around the world, Arnold (1992) compares responses across women with same and mixed sibling sex compositions. He argues that if preferences for a mixed sibling composition exist, then a smaller proportion of women with boys and girls should want another child compared to women with only boys or only girls. However despite having already obtained a mixed sibling sex composition, the majority of women aged 15 to 49 in Sub-Saharan Africa wanted another child (Arnold 1992:94). By contrast, women in Asian, Latin American and North African countries exhibited a clear preference for mixed sibling sex composition. The contrasting results may be explained by desires for larger family sizes in Sub-Saharan African countries. Women may not be concerned about the sex of their first two children if they plan to have many children because eventually a mixed sibling sex composition will be obtained (Ben-Porath and Welch 1976:292). Findings by Arnold (1992) indicate that desires for a mixed sex composition in Sub-Saharan African countries may only begin to emerge among women with four consecutive children of the same sex (*ibid*:94).

Arnold's sample of Sub-Saharan African countries did not include South Africa. Gangadharan and Maitra (2003), however, conduct a specific study of sex preference in South Africa using the 1993 Project for Statistics on Living Standards and Development (PSLSD). They do not find a significant difference in the proportion of African mothers[3] in South Africa (with at least two children) who have a third birth, depending on whether the sex of their first two children was the same or mixed. But they do find that White, Coloured and Indian women in South Africa, with at least two children, are more likely to have third birth if the first two children were of the same sex rather than mixed sex (*ibid*:388). A possible reason for this result is that population groups are heterogeneous in desired sex ratios and family size (Ben-Porath and Welch 1976). Coloureds, Indians and Whites have smaller family sizes than Africans, in which case they may be more concerned about the sibling sex composition of their first two children.

[3] In this study, Gangadharan and Maitra (2003:383) only identify mothers who are *co-resident* with their children using household relationship codes.

5.1 Property i) – exclusion restriction

It is generally difficult to directly test property i), otherwise called the exclusion restriction, because variables captured in the error term cannot be directly observed (Wooldridge 2003:496). However I use basic intuition as well as an analysis of mean contrasts between women with same sex and mixed sibling sex composition to validate the identification strategy.

There are concerns that in some developing countries the exclusion restriction may not hold due to sex-selection, for example through sex-selective abortion or stopping rules.[4] In this case the sex composition of children is non-randomly assigned and may possibly be correlated with background characteristics (Cruces and Galiani 2007:568; Clark 2000:95). Sibling composition variables therefore may not be exogenous within a labour force participation equation. Sex-selection is especially a concern in countries such as India and in North Africa in which strong preferences for sons are exhibited. Son preferences may be driven by tastes and lower net costs of sons compared to daughters (Ben-Porath and Welch 1976). The net cost of a son may be lower because he may be more likely to augment household income and provide old age support (Gangadharan and Maitra 2003:371). Dowries or bride prices may also raise the net cost of girls relative to boys (Clark 2000:95; Ben-Porath and Welch 1976:292). Among Africans in South Africa, however, bridewealth payments, called *ilobolo,* are traditionally made by the potential husband to the bride's family to validate a customary marriage (Casale and Posel 2007). In this case daughters become valuable to their parents for their associated *ilobolo* payments; and therefore the net cost of girls may not exceed the net cost of boys (Ganghadaran and Maitra 2003). For this reason a strong son preference may not exist among Africans in South Africa.

Accordingly, there is no empirical evidence to suggest a strong son preference among African parents in South Africa which could affect the random biological assignment of the

---

[4] Stopping rules, otherwise referred as differential stopping behaviour, occur when couples stop bearing children when they have achieved the desired sex composition. In the case of son preference, couples will continue having children until they reach the desired number of sons (Clark 2000).

sex composition of their offspring. Compared to high infant sex-ratios[5] in countries with strong son preferences, infant sex-ratios in South Africa are low and not skewed in favour of boys (United Nations 2008). Using the 1996 Census, Garenne (2004) reports a sex-ratio of 0.991[6] for the African population under 12 years of age which is even lower than the established average of about 1.05 for European populations. Moreover if strong preferences existed for sons, then average birth intervals between successive births would tend to be shorter if the previous birth was a girl rather than a boy. Among Africans, however, Gangadharan and Maitra (2003) find no significant differences in birth intervals between successive births by the sex of the previous birth.[7] Furthermore, there is no evidence of discrimination against girls in terms of their schooling outcomes. Sibanda and Lehloenya (2005), for example, find that gender differences in school enrolment rates are very small while primary school completion rates are in fact higher among girls than boys.

One way in which to check that the exclusion restriction holds is to compare the average demographic characteristics of women with a mixed sibling sex composition to women with same sex siblings (Aguero and Marks 2008; Angrist and Evans 1998). If there is no correlation between the exogenous instrument and the error term then there should be no systematic differences in the demographic characteristics between women whose first two children are of the same or mixed sex. Table 1 reports the mean differences in demographic characteristics for the two groups of women. Mean differences in age, age at first birth, marital status, rural/urban status and total household earnings do not differ from zero at the five percent level of significance. Specifically, labour force participation rates do not differ across the two groups of women.[8] However compared to co-resident mothers with a mixed sibling sex composition, a larger number of children are born to co-resident mothers whose

---

[5] Sex-ratios are calculated as the number of males to females.

[6] Consistent with the Census 1996, sex-ratios at birth using the Demographic Health Survey (DHS) 1998 and Agincourt data from 1992 to 2000 are reported at 0.995 and 0.997 respectively (Garenne 2004:95).

[7] However they do find evidence of shorter birth intervals after the birth of a daughter among Indians in South Africa suggesting a son preference among this group.

[8] Furthermore results of a basic regression of labour force participation on same sex sibling composition for a combined sample of co-resident and not co-resident African mothers (who have had at least two live births) show no statistically significant effect of sibling composition on labour force participation.

first two children born are of the same sex. Years of schooling are also slightly higher among co-resident mothers who have a mixed sibling sex composition compared to co-resident mothers with a same sex sibling composition, but only at the ten percent level of significance. Furthermore these educational attainment differences are observed only at lower levels of education.[9]

Given the above discussion, there is little evidence that same sex sibling composition violates the exclusion restriction.

5.2 Property ii) – correlation between the instrumental variable and the endogenous explanatory variable

The mean contrasts in Table 1 indicated that among co-resident mothers with at least two children, more children are born to women whose first two children are of the same sex compared to women who have a boy and a girl. Although this suggests the sibling sex composition of a mother's first two children influences fertility outcomes, I test this further using descriptive estimates and first-stage estimation.

[9] It may be argued that women with higher levels of education are more likely to be aware of possible methods to influence the probability of conceiving a boy or a girl by planning intercourse at certain dates within an ovulation cycle. If women have a preference for a mixed sibling sex composition they may be able to exploit these methods to achieve the desired sex composition of children. For this reason the exclusion restriction would be violated because the likelihood of having a mixed sibling sex composition is influenced by background characteristics. But there are two problems with this reasoning. First, if women were to use these methods they would be more educated women with at least a matric or post-matric education. But educational differences across the two groups of women are only observed at primary levels. Furthermore international evidence that couples can influence the sex composition of children conceived by structuring intercourse at different points in the ovulation cycle remains inconclusive (Simpson 1995).

Table 1: Differences in mean characteristics of African mothers by the sex composition of their first two births, 2002

Mean characteristic difference = mean (same sex siblings) – mean (mixed sex siblings)	Co-resident mothers having had at least two births	Not co-resident mothers having had at least two births
Age	-0.251 (0.171) t = -1.466	0.693 (0.457) t = 1.516
Age at first birth	-0.066 (0.098) t = -0.674	0.385 (0.251) t = 1.537
Married	0.003 (0.012) t = 0.218	0.049 (0.032) t = 1.557
Rural	-0.004 (0.012) t = -0.298	-0.015 (0.032) t = -0.471
Total monthly household earnings[a] (2002 prices)	5.538 (76.196) t = 0.0727	-36.383 (146.400) t = -0.2485
Years of schooling	0.195* (0.100) t = 1.937	-0.120 (0.255) t = -0.469
No schooling	-0.017** (0.008) t = -2.203	0.001 (0.020) t = 0.067
Incomplete Primary	-0.013 (0.010) t = -1.308	-0.026 (0.027) t = -0.982
Competed primary	0.014** (0.007) t = 2.006	-0.010 (0.020) t = -0.523
Incomplete Secondary	0.005 (0.012) t = 0.405	0.035 (0.031) t = 1.149
Matric	0.005 (0.009) t = 0.588	-0.007 (0.022) t = -0.328
Post-matric	0.006 (0.006) t = 0.883	0.007 (0.014) t = 0.496
Labour force participant	0.015 (0.012) t = 1.254	-0.030 (0.026) t = -1.174
Number of births	-0.095*** (0.036) t = -2.640	0.0523 (0.083) t = 0.632

Source: GHS 2002

Notes: [a]Total monthly household earnings are in 2002 prices and include only earned income and not unearned income. The data are unweighted. Standard deviations are in parentheses. Means differences are calculated by subtracting the mean characteristic for women with mixed sex siblings from those with same sex siblings. T-statistic is abbreviated by 't'. The null hypothesis of the t-statistic is that the mean difference is zero i.e. no difference exists in the average characteristics of women with a same and mixed sibling sex composition. The sample includes African mothers aged 20 to 49 who have had at least two births. Co-resident mothers are living with at least one of their children aged 18 years or younger. Not co-resident mothers are not living with any of their children. *** 1 % significance level; ** 5% significance level; * 10% significance level.

Table 2 compares the proportion of African women who have had more than two births by the sibling sex composition of their first two children. First, I consider results for a combined or aggregate sample of all mothers, regardless of residency with children and marital status (see first row division in Table 2). In 2002, the proportion of African women having had a third birth was statistically higher among mothers whose first two children were of the same sex rather than different sex. This contrasts with the findings of Gangadharan and Maitra (2003) using the PSLSD 1993.

However preferences for sibling sex composition differ by marital status and co-residency status with children. The 'aggregate' preference among mothers for a mixed sibling sex composition is driven specifically by the sample of co-resident mothers who are unmarried. Among this group, who constitute 34 percent of the total sample of African mothers with at least two births, the proportion having had a third birth is statistically higher among mothers whose first two children were of the same sex rather than opposite sex. However, among married co-resident mothers and all mothers not co-resident with their children, there are no significant mean contrasts in the proportion having had a third birth by sibling sex composition. I however acknowledge that insignificant mean contrasts among the sample of not co-resident mothers may be due to small sample sizes.

In addition to descriptive statistics, first-stage estimations can be used to directly test property ii) – whether a correlation exists between the endogenous explanatory variable and the instrument. As specified in equations (2a) and (2b), I estimate the relationship between more than two births and sibling sex composition. The results are presented in Table 3.

Table 2: Proportion of mothers who have had a third birth, African mothers 2002

	All mothers (not distinguished by co-residency with children)		
Sex of first two children among women with two or more children	**Married and unmarried**	**Married**	**Unmarried**
(1) Mixed sibling sex	0.602 (0.490) N = 2,223	0.682 (0.466) N = 1,497	0.484 (0.500) N = 726
(2) Same sex sibling	0.620 (0.485) N = 2,363	0.697 (0.460) N = 1,547	0.513 (0.500) N = 816
Two girls	0.623 (0.485) N = 1,186	0.698 (0.459) N = 797	0.525 (0.500) N = 389
Two boys	0.622 (0.485) N = 1,191	0.702 (0.457) N = 763	0.504 (0.500) N = 428
Mean difference (2) – (1)	0.019** (0.011) t = 1.6530	0.016 (0.014) t = 1.1132	0.029* (0.018) t = 1.601
		Co-resident mothers	
Sex of first two children among women with two or more children	**Married and unmarried**	**Married**	**Unmarried**
(3) Mixed sibling sex composition	0.617 (0.486) N = 1,990	0.694 (0.461) N = 1,369	0.496 (0.500) N = 621
(4) Same sex sibling composition	0.638 (0.481) N = 2,099	0.710 (0.453) N = 1,421	0.526 (0.500) N = 678
Two boys	0.635 (0.481) N = 1,048	0.711 (0.454) N = 728	0.512 (0.500) N = 320
Two girls	0.646 (0.478) N = 1,065	0.716 (0.451) N = 706	0.543 (0.500) N = 359
Mean Difference (4) – (3)	0.021** (0.012) t = 1.7373	0.017 (0.015) t = 1.1525	0.030* (0.020) t = 1.4933
		Not co-resident mothers	
Sex of first two children among women with two or more children	**Married and unmarried**	**Married**	**Unmarried**
(5) Mixed sibling sex composition	0.496 (0.501) N = 233	0.577 (0.495) N = 128	0.424 (0.495) N = 105
(6) Same sex sibling composition	0.510 (0.500) N = 264	0.580 (0.495) N = 126	0.458 (0.499) N = 138
Two boys	0.535 (0.500) N = 126	0.619 (0.488) N = 57	0.469 (0.501) N = 69
Two girls	0.485 (0.501) N = 138	0.538 (0.501) N = 69	0.448 (0.499) N = 69
Mean Difference (6) – (5)	0.014 (0.032) t = 0.4354	0.003 (0.047) t = 0.0707	0.034 (0.043) T = 0.8100

Source: GHS 2002

Notes: Standard deviations are in parentheses. Data are unweighted. Sample is of African mothers aged 20 to 49 who have had *at least two births*. A co-resident mother is a woman co-residing with at least one of her own children aged 18 years or younger. A not co-resident mother is a woman with own children aged 18 years or younger but not living with any of her children. The alternative hypothesis is that the mean difference is greater than zero. *** 1 % significance level; ** 5% significance level; * 10% significance level.

Even after partialling out the effects of other explanatory variables included in the reduced form equation, the first-stage results in Table 3 are generally consistent with the descriptive results in Table 2. Considering the combined sample of all mothers (not distinguished by residency with children and marital status), the aggregate coefficient on same sex siblings is positive and significant at the five percent level but small in magnitude. Compared to mothers with a mixed sibling sex composition, mothers with a same sex sibling composition are only about two percent more likely to have a third birth. The results also indicate that on average there is a bias for boys: the coefficient on two girls is larger and more significant than the coefficient on two boys. Therefore mothers whose first two children are girls rather than boys are more likely to have a third birth in the hope of conceiving a boy.

Again the aggregate results are driven *only* by the sample of *unmarried* mothers who are co-resident with at least one of their children. The coefficients on sibling sex composition variables are almost always insignificant for the sample of married[10] co-resident mothers and for both married and unmarried not co-resident mothers. Therefore heterogeneous preferences exist for desired sibling sex composition among the different samples of mothers. This may be attributed to differences in family size across the different groups. Table 4 shows that compared to the group of married co-resident mothers, unmarried co-resident mothers have on average fewer children and may therefore be more concerned about the sex composition of their first few children.

[10] It may be criticised that statistically insignificant coefficients are obtained on sibling sex variables for the sample of married mothers because included in the sample are married mothers who are not living with their spouse or partner. Married parents who do not live together may be less likely to form preferences for family size and composition than married parents who live apart. But I find that results are robust to the inclusion or exclusion of married mothers who are not living with their spouse or partner in the sample of married women.

Table 3: First-stage results - dependent variable is *more than two births,* T_i

	All mothers (not distinguished by co-residency with children)		
	Married and unmarried	**Married**	**Unmarried**
Instrument: Same sex siblings			
Coefficient – same sex	0.020** (0.009)	0.011 (0.012)	0.031** (0.015)
Partial R^2	0.0006	0.0002	0.0014
F-statistic	4.541	0.822	4.216
Instrument: Two girls, two boys			
Coefficient – Two Girls	0.023** (0.011)	0.010 (0.014)	0.043** (0.018)
Coefficient – Two Boys	0.016 (0.011)	0.011 (0.014)	0.019 (0.019)
Partial R^2	0.0006	0.0002	0.0018
F-statistic	2.406	0.384	2.751
N	7,477	4,389	3,088

	Co-resident mothers		
	Married and unmarried	**Married**	**Unmarried**
Instrument: Same sex siblings			
Coefficient – same sex	0.018* (0.010)	0.011 (0.012)	0.026* (0.017)
Partial R^2	0.0005	0.0002	0.0010
F-statistic	3.24202	0.840	2.499
Instrument: Two girls, two boys			
Coefficient – Two Girls	0.026** (0.012)	0.013 (0.015)	0.045** (0.020)
Coefficient – Two Boys	0.009 (0.012)	0.008 (0.015)	0.008 (0.021)
Partial R^2	0.0007	0.0002	0.0020
F-statistic	2.303	0.419	2.508
N	6,487	3,949	2,538

	Not co-resident mothers		
	Married and unmarried	**Married**	**Unmarried**
Instrument: Same sex siblings			
Coefficient – same sex	0.039 (0.028)	0.015 (0.041)	0.056 (0.037)
Partial R^2	0.0021	0.0003	0.0044
F-statistic	2.019	0.128	2.30815
Instrument: Two girls, two boys			
Coefficient – Two Girls	0.017 (0.034)	-0.012 (0.051)	0.048 (0.045)
Coefficient – Two Boys	0.062* (0.034)	0.041 (0.051)	0.065 (0.045)
Partial R^2	0.0035	0.0022	0.0046
F-statistic	1.674	0.452	1.203
N	990	440	550

Source: GHS 2002

Notes: Data are unweighted. Standard deviations are in parentheses. Sample is of African mothers aged 20 to 49 who have had *at least two births*. A co-resident mother is a woman co-residing with at least one of her own children aged 18 years or younger. A not co-resident mother is a woman with own children aged 18 years or younger but not living with any of her children. Other exogenous variables include the woman's age, age squared, age at first birth, her educational status and marital status. Controls are also included for household compositional variables, dummy variables for province of residence and whether the woman lives in a rural or urban area. *** 1 % significance level; ** 5% significance level; * 10% significance level.

An interesting result from the first-stage estimations is the bias for boys among unmarried mothers who are co-resident with at least one of their own children. This bias emerges despite a context of bridewealth payments which raises the value of daughters. A possible explanation is that in the absence of a spouse/partner to provide financial support, unmarried co-resident mothers may have a preference for sons who may be better positioned in the market place than daughters to support their mothers in old age. But no son preference is evident among the sample of mothers who are unmarried but are not co-residing with any of their children. One reason for this is that they may be more likely to be working, therefore more economically independent. This may mitigate the need for financial support from a son. Table 4 indicates that 84 percent of unmarried not co-resident mothers (with at least two children) were labour force participants compared to only 69 percent of unmarried co-resident mothers (with at least two children).

Table 3: Number of live births and labour force participation rates among African mothers having had at least two births, 2002

	Co-resident mothers		**Not co-resident mother**	
	Married	**Unmarried**	**Married**	**Unmarried**
Number of births	3.631 (1.559)	2.986 (1.275)	3.259 (1.437)	2.803 (1.136)
Labour force participation	0.594 (0.491)	0.685 (0.464)	0.736 (0.441)	0.837 (0.370)
N	3,949	2,538	440	550

Source: GHS 2002
Notes: Data are unweighted. Standard deviations are in parentheses. Sample is of African mothers aged 20 to 49 who have had *at least two births*. A co-resident mother is a woman co-residing with at least one of her own children aged 18 years or younger. A not co-resident mother is a woman with own children aged 18 years or younger but not living with any of her children.

Despite a statistically significant relationship between same sibling sex composition and having a third birth among the aggregate sample of mothers, this result is driven only by a select sample of mothers. Sibling sex composition is a poor instrument for further childbearing among married co-resident mothers and all not co-resident mothers. Furthermore, the quality of the instrumental variable is poor given the small values of the partial R^2 and F-statistic of the identifying instruments in the first-stage estimation (Bound *et al* 1995). This result holds regardless of the sample used. The partial R^2 values measure the strength of the linear relationship between the endogenous explanatory variable and the identifying instrument once the effects of other exogenous variables have been partialled out (Wooldridge 2003:490, Bound *et al* 1995:444). The partial R^2 value of the identifying instruments, same sex sibling composition or two girls and two boys, is less than 0.005 for every sample identified in Table 3. Moreover, F-statistics for the single instrument, same sex siblings, always lie well below the suggested value of ten for a strong instrument (Stock *et al* 2002:522). For this reason I conclude that sibling sex composition is a weak instrument for further childbearing among African mothers in South Africa, regardless of co-residency with children or marital status.

As a consequence of weak instruments, the IV estimates of the effects of childbearing on labour force participation are likely to be more unreliable than OLS estimates. I illustrate this in the next section.

6. Ordinary least squares and second-stage results

Table 5 presents second-stage IV estimates using same sex sibling composition or two girls and two boys, as instruments for a third birth. It also presents OLS estimates for comparative purposes. The table splits the sample according to residency with children and marital status. Considering the aggregate sample of mothers (not distinguished by residency with children and marital status) OLS results show that having a third birth does not have a statistically significant effect on a mother's labour force participation. When disaggregated by marital status, the OLS estimate on *more than two births* is significant but differs in sign depending on marital status. If mothers have a third birth they are significantly less likely to be labour force participants if they are married, but if they are unmarried they are more likely to be labour force participants. There is no effect of having a third birth on the labour force participation of not co-resident mothers.

IV estimates differ considerably from OLS estimates; they are ‘very noisy’ and are estimated with little precision. Moreover IV estimates provide evidence of implausible coefficients. For example, IV estimates using same sex siblings in the second row division of Table 5 indicate that among unmarried co-resident mothers, having a third child *lowers* their labour force participation by as much as 93 percent. This estimate lies within a 95 percent confidence interval ranging from negative 268 percent to a positive value of 83 percent. The IV estimate therefore cannot be interpreted with any level of confidence. Among the same group of women, the OLS estimate indicates that they are five percent *more* likely to be labour force participants if they have a third child. This OLS estimate can be interpreted with more precision than the IV estimate because its 95 percent confidence interval is considerably smaller, ranging from one percent to nine percent (confidence intervals are not presented in the table).

The noisy and imprecise IV estimates in Table 5 can be attributed to the weak explanatory power of sibling sex composition variables in the first-stage estimations; and the cost of this weak explanatory power is very large IV standard errors in the second-stage. Standard errors of IV estimates are always expected to be greater than those of OLS estimates due to imperfect correlation between the endogenous explanatory variable and the identifying instrument (Wooldridge 2003). Consider the equation below for the asymptotic variance (or square of the asymptotic standard error) of an IV estimate,

$$\sigma^2 / (STT_T{*}R^2)$$

where SST_T is the total sum of squares of T_i, σ^2 is the variance of e_i and R^2 measures the strength of the linear relationship between the having a third child and the identifying instrument, same sex sibling composition. An R^2 value of less than one will always raise the IV standard error above that of the OLS standard error (Wooldridge 2003:490). In this case partial R^2 values in first-stage estimations are always less than 0.005 (see Table 3). It is therefore not surprising that IV standard errors are 20 to 200 times greater than the standard errors of OLS estimates. For example, among not co-resident mothers who are married, the standard error on the OLS estimate is 0.05 while the standard error on the IV estimate using same sex siblings is 200 times larger at 10.23 (see third row division in Table 5).

Table 4: Second-stage and OLS results - dependent variable is *labour force participation*, Y_i

	All mothers (not distinguished by co-residency with children)		
	Married and unmarried	**Married**	**Unmarried**
OLS	-0.013 (0.013)	-0.051*** (0.018)	0.037** (0.018)
IV: same sex	-0.230 (0.525)	0.599 (1.481)	-0.546 (0.558)
IV: two girls and two boys	-0.038 (0.501)	0.346 (1.415)	-0.083 (0.424)
N	7,477	4,389	3,088
	Co-resident mothers		
	Married and unmarried	**Married**	**Unmarried**
OLS	-0.007 (0.014)	- 0.048** (0.019)	0.048** (0.021)
IV: same sex	-0.565 (0.700)	0.099 (1.316)	-0.928 (0.895)
IV: two girls and two boys	0.032 (0.527)	0.145 (1.324)	-0.023 (0.459)
N	6,487	3,949	2,538
	Not co-resident mothers		
	Married and unmarried	**Married**	**Unmarried**
OLS	0.016 (0.028)	-0.020 (0.050)	0.030 (0.034)
IV: same sex	0.631 (0.757)	3.601 (10.233)	-0.003 (0.500)
IV: two girls and two boys	0.153 (0.487)	0.306 (1.120)	-0.080 (0.493)
N	990	440	550

Source: GHS 2002

Notes: Data are unweighted. Standard errors are in parentheses. Sample is of African mothers aged 20 to 49 who have had *at least two births*. A co-resident mother is a woman co-residing with at least one of her own children aged 18 years or younger. A not co-resident mother is a woman with own children aged 18 years or younger but not living with any of her children. Other exogenous variables include the woman's age, age squared, age at first birth, her educational status and marital status. Controls are also included for household compositional variables, dummy variables for province of residence and whether the woman lives in a rural or urban area. *** 1 % significance level; ** 5% significance level; * 10% significance level.

Due to large standard errors, IV estimates cannot be interpreted with any level of confidence and are meaningless (Wooldridge 2003:494). In the event that estimations could be run with larger sample sizes, reducing the standard errors of IV estimates, the weak correlation between further childbearing and sibling sex composition variables would result in IV estimates that are biased in the same direction as OLS estimates (Bound *et al* 1995:443).

Conclusion

International studies argue that same sex sibling composition is a good instrument for further childbearing, satisfying the two properties of an instrumental variable (Cruces and Galiani 2007; Iacovou 2001; Angrist and Evans 1998). Consistent with these studies there is little evidence that sibling sex composition among Africans in South Africa violates property i) (or the exclusion restriction). There is no evidence of strong sex-selection among African mothers in South Africa which would affect the random biological assignment of a child's sex. This study also finds no systematic differences in background characteristics across mothers by the sibling sex composition of their first two children.

However this study reveals that among African women sibling sex composition does not satisfy the second IV property – whether there is a strong correlation between the endogenous explanatory variable and the instrument. Among African mothers, same sex sibling composition has weak explanatory power in first-stage estimations of the propensity to have more children. Where correlations were observed between sibling sex composition variables and further childbearing, the result was driven only by a select sample of unmarried mothers who are co-resident with their children. Furthermore partial R^2 values and F-statistics are very low in first-stage estimations, regardless of residency with children or marital status.

The poor explanatory power of the sibling sex composition instruments in first-stage estimations translates into very large standard errors of IV estimates in the second-stage. In this study, standard errors on IV estimates are 20 to 200 times larger than standard errors of OLS estimates. Compared to OLS estimates, the resulting IV estimates are 'very noisy', estimated with little precision and have no meaningful interpretation.

Same sex sibling composition is a weak instrument for childbearing among African women in South Africa. A possible reason for this result is that populations are heterogeneous in desired sex-ratios and family size (Ben-Porath and Welch 1976:292). Among Africans in South Africa family sizes are generally larger than in the other countries in which same sex sibling composition is identified as a good instrument for childbearing (Central Intelligence Agency 2008). With larger desired family sizes, African women may be less concerned about the sex of their first two children because they expect to achieve a mixed sibling sex composition when more children are born (Ben-Porath and Welch 1976). This hypothesis may also explain differences in observed correlations between further childbearing and sibling sex composition across the different samples of African mothers, disaggregated by co-residency with children and marital status.

References

Aguero, J. and Marks, M. (2008) “Motherhood and Female Labor Force Participation: Evidence from Infertility Shocks”. *American Economic Review* 98(2):500-504.

Angrist, J. and Evans, W. (1998) "Children and their parents' labour supply: Evidence from exogenous variation in family size". *The American Economic Review* 88(3):450-477.

Arnold, F. (1992) “Sex Preferences and Its Demographic and Health Implications”. *International Family Planning Perspectives* 18(3):93-101.

Ben-Porath, Y. and Welch, F. (1976) “Do sex preferences really matter?” *The Quarterly Journal of Economics* 90(2):285-307.

Bound, J., Jaeger, D. and Baker, R. (1995) “Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous

Explanatory Variable is Weak”. *Journal of the American Statistical Association* 90(430):443-450.

Boushey, H. (2008) “‘Opting Out?’ The Effect of Children on Women’s Employment in the United States”. *Feminist Economics* 14(1):1 – 36.

Browning, M. (1992) “Children and Household Economic Behaviour”. *Journal of Economic Literature* 30(3):1434-1475.

Bronars, G. and Grogger, J. (1994) The economic consequences of unwed motherhood: Using twin births as a natural experiment. *The American Economic Review* 84(5):1141 – 1156.

Casale, D. and Posel, D. (2007) “The Male Marital Earnings Premium in the Context of Bridewealth Payments: Evidence from South Africa”. ERSA Working Paper, no. 57.

Chapman, B., Dunlop, Y., Gray, M., Liu, A. and Mitchell, D. (2001) "The impact of children on the earnings lifetime of Australian women: Evidence from the 1990’s". *The Australian Economic Review* 34(4): 373 – 389.

Central Intelligence Agency (2008) “World Fact Book: Rank Order – Total Fertility Rate”. (Accessed 10 September 2008). Available: https://www.cia.gov/library/publications/the-world-factbook/rankorder/2127rank.html

Clark, S. (2000) “Son Preference and Sex Composition of Children: Evidence from India”. *Demography* 37(1):95-108.

Connelly, R., DeGraff, D., Levinson, D. and McCall, B. (2006) “Tackling the Endogeneity of Fertility in the Study of Women’s Employment in Developing Countries: Alternative Estimation Strategies Using Data From Urban Brazil”. *Feminist Economics* 12(4):561-597.

Cruces, G. and Galiani, S. (2007) "Fertility and female labor supply in Latin America: New causal evidence". *Labour Economics* 14(3):565 – 573.

Gangadharan, and Maitra, P. (2003) “Testing for Son Preference in South Africa”. *Journal of African Economics* 12(3):371 – 416.

Garenne, M. (2004) “Sex ratios at birth in populations of Eastern and Southern Africa”. *Southern African Journal of Demography* 9(1):91-96.

Gormick, J., Meyers, M. and Ross, K. (1996) “Public polices and employment of mothers: A cross-national study”. *Luxembourg Income Study Working Paper Series*, Working paper no. 140.

Iacovou, M. (2001) “Fertility and Female Labour Supply”. *Working Papers of the Institute for Social and Economic Research*, paper 2001-19. Colchester: University of Essex.

Posel, D. and van der Stoep, G. (2008) “Co-resident and Absent mothers: Motherhood and Labour Force Participation in South Africa”. Paper presented at the AES conference, Pretoria, July 2008 and at the Conference on Income Distribution and the Family, Kiel (Germany).

Rosenzweig, M. and Wolpin, K. (1980) "Life-cycle labour supply and fertility: Causal inferences from household models". *The Journal of Political Economy* 88(2): 328 – 348.

Sibanda, A. and Lehloenya, G. (2005) “Race and Gender Gaps in Education”. In: Zuberi, T., Sibanda, A. and Udjo, E. (eds) (2005) *The Demography of South Africa*. M.E. Sharpe Inc. pp. 253 – 265.

Simpson, J. (1995) “Pregnancy and the timing of intercourse”. The *New England Journal of Medicince* 333 (23):1563-1565.

Statistics South Africa, *General Household Survey*, 2002. Pretoria

Stock, J., Wright, J. and Yogo, M. (2002) “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments”. *Journal of Business and Economic Statistics* 20(4):518-529.

United Nations (2008) “World Population Prospects: The 2006 Revision – population sex ratios”. (Accessed 9 September 2008) Available: http://data.un.org

Vere, J. (2008) “Dragon Children: Identifying the Causal Effect of the First Child on Female Labour Supply with the Chinese Lunar Calendar”. *The Oxford Bulletin of Economics and Statistics* 70(3):303-325.

Wooldridge, J. (2003) “Chapter 15: Instrumental Variables Estimation and Two-stage Least Squares”. In: *Introductory Econometrics, A Modern Approach, second edition.* South-Western Publishing.